MAIL STOP APPEAL BRIEF - PATENTS

TRANSMITTAL OF APPEAL BRIEF (Large Entity)	Docket No. PNET.009D

In Re Application Of: **Tsukasa Yajima**

Serial No.	Filing Date	Examiner	Group Art Unit
09/768,271	January 25, 2001	A. Mai	2814

Invention: **SEMICONDUCTOR DEVICE HAVING PROTECTIVE LAYER ON FIELD OXIDE**

TO THE COMMISSIONER FOR PATENTS:

Transmitted herewith in triplicate is the Appeal Brief in this application, with respect to the Notice of Appeal filed on **July 9, 2003.**

The fee for filing this Appeal Brief is: **$320.00**

☒ A check in the amount of the fee is enclosed.

☐ The Director has already been authorized to charge fees in this application to a Deposit Account.

☒ The Director is hereby authorized to charge any fees which may be required, or credit any overpayment to Deposit Account No. **50-0238**

Signature

Dated: **September 5, 2003**

ANDREW J. TELESZ, JR.
REG. NO. 33,581

VOLENTINE FRANCOS, P.L.L.C.
12200 SUNRISE VALLEY DRIVE, SUITE 150
RESTON, VA 20191

TEL. NO. (703) 715-0870

cc:

I certify that this document and fee is being deposited on with the U.S. Postal Service as first class mail under 37 C.F.R. 1.8 and is addressed to the Commissioner for Patents, P.O. Box 1450, Alexandria, VA 22313-1450.

Signature of Person Mailing Correspondence

Typed or Printed Name of Person Mailing Correspondence

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

In re Patent Application of	:	Before the Board of Appeals
Tsukasa Yajima	:	Appeal No.:
Serial No.: 09/768,271	:	Group No.: 2814
Filed: January 25, 2001	:	Examiner: A. Mai

For: SEMICONDUCTOR DEVICE HAVING PROTECTIVE LAYER ON FIELD OXIDE

September 5, 2003

APPEAL BRIEF

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

In re Patent Application of	:	Before the Board of Appeals
Tsukasa Yajima	:	Appeal No.:
Serial No.: 09/768,271	:	Group No.: 2814
Filed: January 25, 2001	:	Examiner: A. Mai

For: SEMICONDUCTOR DEVICE HAVING PROTECTIVE LAYER ON FIELD OXIDE

September 5, 2003

TABLE OF CONTENTS

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

In re Patent Application of	:	Before the Board of Appeals
Tsukasa Yajima	:	Appeal No.:
Serial No.: 09/768,271	:	Group No.: 2814
Filed: January 25, 2001	:	Examiner: A. Mai

For: SEMICONDUCTOR DEVICE HAVING PROTECTIVE LAYER ON FIELD OXIDE

APPEAL BRIEF

U.S. Patents and Trademark Office
2011 South Clark Place
Customer Window, **Mail Stop Appeal Brief-Patents**
Crystal Plaza Two, Lobby, Room 1B03
Arlington, VA 22202

Date: September 5, 2003

Sir:

This is an appeal from the final rejection of claims 6-9 and 11-19, which claims were finally rejected in the Office Action dated January 10, 2003. A Notice of Appeal was filed on July 9, 2003.

I. REAL PARTY IN INTEREST

This application is assigned to Oki Electric Industry Co., Ltd., which is the real party in interest.

II. RELATED APPEALS AND INTERFERENCES

There are no other appeals or interferences that will directly affect or be affected by or have a bearing on the Board's decision in this pending appeal.

III. STATUS OF THE CLAIMS

Claims 6-9 and 11-19 are present in this application, and the rejections thereof are hereby appealed.

IV. STATUS OF AMENDMENTS

Subsequent to the Final Office Action dated January 10, 2003, Appellant submitted an Amendment under 37 C.F.R. 1.116 on May 12, 2003, canceling proposed Fig. 1(i); amending the paragraph beginning on page 13, line 12 of the specification; and amending claims 6, 11 and 16.

In the Advisory Action dated June 3, 2003, the Examiner indicated that the Amendment filed on May 12, 2003, had been entered. Thus, the appended copy of the claims involved in this Appeal, claims 6-9 and 11-19, reflect the amendments made and entered via the Amendment under 37 C.F.R. 1.116 filed on May 12, 2003.

V. SUMMARY OF THE INVENTION

As illustrated in Fig. 1(a), the semiconductor device of an embodiment of the present application includes a silicon oxide layer 32 of 15-20 nm thickness on silicon

substrate 31, and a silicon nitride layer 33 of 400-600 nm thickness as an oxidation resisting layer on silicon oxide layer 32 (page 10, lines 13-18). After an aperture is formed in silicon nitride layer 33 and silicon oxide layer 32 as illustrated in Fig. 1(b), ions are implanted into silicon substrate 31 through the aperture to prevent formation of a parasitic transistor (page 10, line 22 through to page 11, line 4). Thereafter, the silicon substrate 31 is wet-oxidized to form field oxide 34 having thickness of about 1 μm in the aperture, as illustrated in Fig. 1(c) (page 11, lines 5-8).

As illustrated in Fig. 1(d), a polysilicon layer 11 having a thickness of about 1 μm is deposited as a protection layer on field oxide 34 within the aperture and on silicon nitride layer 33 (page 11, lines 11-12). Polysilicon layer 11 is then polished such as by chemical mechanical polishing (CMP), until the surface of silicon nitride layer 33 is exposed and so that polysilicon layer 12 remains in the aperture, as illustrated in Fig. 1(e) (page 11, lines 14-17). The exposed silicon nitride layer 33 is then removed by wet chemical etching, so that polysilicon layer 12 remains as a protective layer formed on only field oxide 34, as illustrated in Fig. 1(f) (page 12, line 21 through to page 13, line 3).

Thereafter, photolithography and etching are carried out to form gate oxide layer 35a, polysilicon layer 35b for gate electrodes, and tungsten silicide 35c, to form gates 35, as illustrated in Fig. 1(g) (page 13, lines 3-8). Oxide layer 36 for subsequent formation of side-walls, is then deposited over the entire surface of the structure including gates 35, silicon substrate 31, field oxide 34 and polysilicon layer 12, as also

illustrated in Fig. 1(g) (page 13, lines 9-11).

Thereafter, oxide layer 36 is etched by reactive ion etching to form side-walls 37 on gates 35, as illustrated in Fig. 1(h) (page 13, lines 12-14). However, since non-uniformities in etching speed and of oxide layer 36 thickness exist, overetching is carried out to completely remove oxide layer 36 from the top of gates 35 and from silicon substrate 31. Polysilicon layer 12 prevents etching of field oxide 34 during this overetching of oxide layer 36, preventing decreases of field isolation voltage caused by thinning of field oxide 34 (page 14, lines 1-7). The MOSFETs are then completed by forming an insulating layer on the structure, forming contact holes in the insulating layer, and depositing aluminum wiring (not shown) (page 13, lines 14-17).

VI. ISSUES

Claims 6-9 and 11-19 stand rejected under 35 U.S.C. 102(b) as being clearly anticipated by the Yoo et al. reference (U.S. Patent No. 5,605,853).

Accordingly, one issue presented is whether or not the semiconductor devices of claims 6-9 and 11-19 are anticipated by the teaching of the Yoo et al. reference.

Appellant emphasizes that the objection to the drawings under 37 C.F.R. 1.83(a) and the objection to the Amendment filed on October 31, 2002, under 35 U.S.C. 132, as stated on page 2 of the Final Office Action dated January 10, 2003, have been withdrawn as acknowledged in the Advisory Action dated June 3, 2003. Accordingly, the objection to the drawings under 37 C.F.R. 1.83(a) and the objection to the

Amendment under 35 U.S.C. 132 are not issues of consideration.

VII. GROUPING OF CLAIMS

Appellant respectfully wish to group all of claims 6-9 and 11-19 together.

VIII. ARGUMENTS

A. Rejection of Claims 6-9 and 11-19 under 35 U.S.C. 102(b) as Being Clearly Anticipated by the Yoo et al. Reference

The semiconductor device of claim 6 includes in combination a field oxide "formed on the substrate between said first and second gates" and a protective layer "formed selectively on said field oxide to prevent overetching of said field oxide, said protective layer being a material different than said field oxide".

On page 3 of the Final Office Action dated January 10, 2003, the Examiner has alleged that floating gate 21 as illustrated in Fig. 2 of the Yoo et al. reference may be interpreted as the protective layer of claim 6. However, Appellant respectfully submits that floating gate 21 in Figs. 2-7 of the Yoo et al. reference cannot be interpreted as the protective layer of claim 6, because floating gate 21 cannot be formed on field oxide (FOX) layer 12 (as would be necessary to meet the features of claim 6) and function as a working floating gate.

It should be understood that to function, a simple floating gate in theory must necessarily exchange charges with a diffusion layer formed on the surface of a

semiconductor substrate. Thus, a floating gate typically must be formed on a relatively thin insulating layer, such as a tunnel oxide layer, so that exchange of charges between the diffusion layer and the floating gate may occur. Floating gates generally are not formed on a FOX layer, because exchange of charges between a floating gate and a corresponding semiconductor substrate through a relatively thick FOX layer cannot occur.

In order to emphasize this point, an I.E.E.E. Electron Device Letters publication by Haddad et al. entitled "Degradations Due to Hole Trapping In Flash Memory Cells" was submitted as of interest along with the Amendment dated May 12, 2003. (Copies of the I.E.E.E. Electron Device Letters publication in triplicate are provided in Appendix B.) The I.E.E.E. Electron Device Letters publication specifically concerns flash memory cells. As described in the bottom paragraph in the first column on page 117 of the publication, the gate oxide used is less than 120Å (12nm) thick and the erasure mechanism is dominated by Fowler-Nordheim (F-N) tunneling. Accordingly, it should be understood by one of ordinary skill that in order to perform writing and deletion by means of F-N tunneling in flash memory cells, it is necessary to use a gate oxide having a thickness of approximately 120Å. Particularly, if the gate oxide is too thick, it is impossible to inject and emit electrons through the gate oxide into the floating gate by means of F-N tunneling.

To further emphasize this point, the Mizutani patent (U.S. Patent No. 4,637,128) was also submitted as of interest along with the Amendment dated May 12, 2003.

(Copies of the Mitzutani patent in triplicate are also provided in Appendix B.) As described beginning in column 4, line 57 of the Mitzutani patent, after the ion implantation step, a field insulation layer 60 having a thickness of 6000Å (600nm) is formed, as shown in Fig. 4D. In contrast, as described in column 5, lines 14-18 of the Mitzutani patent, a thermal oxide film 66 is formed to be so thin so as to be generally removed in the patterning step of the poly-Si layer, but is not removed at the portion below the floating gate electrode 68. That is, a relatively thin oxide layer is formed under the floating gate, as opposed to the relatively thicker field oxide layer. Thus, it should be understood that floating gate electrodes in general are formed on relatively thin gate oxides, as opposed to relatively thick field oxide layers.

To still further emphasize this point, an excerpt from the Wolf text Silicon Processing for the VLSI Era, Volume 2: Process Integration is newly submitted herewith as of interest. (Copies of the excerpt of the Wolf text in triplicate are also provided in Appendix B.) As illustrated in Fig. 8-33 on page 624 of the excerpt, the charges are transferred from the silicon substrate to the floating polysilicon gate through the relatively thin gate oxide. As further described with respect to floating-gate tunneling oxide "FLOTOX" transistors with reference to Fig. 8-38 as beginning on page 629 of the excerpt, a thin gate oxide of 8-12nm is used near the drain of the structure, and the remainder of the floating-gate oxide is typically 50nm thick. Programming of the FLOTOX transistor is done by causing electrons to be transferred from the substrate to the floating gate through the thin oxide layer by means of F-N tunneling. Thus, it should

be understood that floating gates in general are formed on relatively thin gate oxides, as opposed to relatively thick field oxide layers.

In the Final Office Action dated January 10, 2003, the Examiner asserted in the Response to Argument section on page 7 thereof, that one of ordinary skill should conclude "without any reasonably doubt" that layer 21 of the Yoo et al. reference is formed on FOX layer 12. However, as evidenced by the I.E.E.E. Electron Device Letters publication and the Mitzutani patent subsequently submitted as of interest in the Amendment dated May 12, 2003, and the Wolf excerpt as presented herewith, functional floating gates are typically formed on relatively thin gate oxides of approximately 120Å (12nm), not relatively thick field oxides. Contrary to the Examiner's assertion, one of ordinary skill should understand that floating gate 21 in Figs. 2-7 of the Yoo et al. reference cannot be formed on FOX layer 12 and be a functional floating gate. It consequently follows that it cannot be concluded "without any reasonable doubt" that layer 21 of the Yoo et al. reference is formed on FOX layer 12, as asserted by the Examiner.

Particularly, the floating gate of the Yoo et al. reference is described in column 3, lines 15-17 as formed **over** a field oxide region, not specifically on a field oxide region. More particularly, as further described in column 4, lines 2-4 of the Yoo et al. reference "Also formed is floating gate 21 for the floating gate memory cell (for an EEPROM) **in region 70**" (our emphasis added). Floating gate 21 of the Yoo et al. reference is not specifically described as formed on FOX layer 12.

In the Advisory Action dated June 3, 2003, in consideration of the arguments presented in the Amendment dated May 12, 2003, the Examiner has noted on page 2 thereof that "since the layer 21 of Yoo merely formed on the Fox layer 12, as shown in Fig. 2, it meets the limitation of the claim, e.g. a protective layer formed on said field oxide". However, this general assertion by the Examiner regarding floating gate 21 of the Yoo et al. reference appears contrary to the above noted teachings of the I.E.E.E. Electron Device Letters publication and the Mitzutani patent regarding floating gates and gate oxide thickness, and as also contrary to the teaching in the Wolf excerpt. In absence of any additionally relied upon teaching or reasoning as offered by the Examiner showing that a functional floating gate can be formed on a relatively thick field oxide layer, the Examiner has failed to clearly answer the substance of the traversal and has failed to clearly develop this issue as now on Appeal.

Accordingly, Appellant respectfully submits that because floating gate 21 of the Yoo et al. reference cannot be formed on FOX layer 12 and function as a working floating gate, floating gate 21 of the Yoo et al. reference cannot be interpreted as the protective layer of claim 6. Appellant therefore respectfully submits that the semiconductor device of claim 6 distinguishes over the Yoo et al. reference as relied upon by the Examiner, and that this rejection of claims 6-9 is therefore improper for at least these reasons.

The semiconductor device of claim 11 includes in combination a field oxide "formed on the substrate adjacent the active region", and a protective layer "formed on

said field oxide to prevent overetching of said field oxide, said protective layer being a material different than said field oxide...said protective layer being formed on said field oxide only". The semiconductor device of claim 16 similarly includes in combination a field oxide "formed on the substrate adjacent the active region", and a protective layer "formed on said field oxide to prevent overetching of said field oxide, said protective layer being a material different than said field oxide...said protective layer being formed on said field oxide only".

Appellant respectfully submits that floating gate 21 of the Yoo et al. reference cannot be interpreted as the protective layer of respective claims 11 and 16, for at least similar reasons as set forth previously with respect to claim 6. Particularly, since floating gate 21 of the Yoo et al. reference cannot be formed on FOX layer 12 and function as a working floating gate, floating gate 21 of the Yoo et al. reference cannot be interpreted as the protective layer of respective claims 11 and 16. Appellant therefore respectfully submits that the semiconductor devices of respective claims 11 and 16 distinguish over the Yoo et al. reference as relied upon by the Examiner, and that this rejection of claims 11-19 is improper for at least these reasons.

B. Conclusion

Accordingly, it is respectively submitted that the rejection of claims 6-9 and 11-19 as being clearly anticipated by the Yoo et al. reference is erroneous and should be reversed.

Consequently, favorable reconsideration and allowance of all claims by the Honorable Board of Patent Appeals and Interferences is respectfully requested.

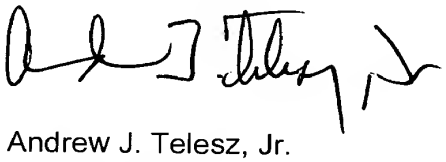
In the event that there are any outstanding matters remaining in the present application, please contact Andrew J. Telesz, Jr. (Reg. No. 33,581) at (703) 715-0870 in the Washington, D.C. area, to discuss these matters.

The required fee of $320.00 under 37 C.F.R. 1.17(c) for filing this Appeal Brief is attached hereto.

If necessary, the Commissioner is hereby authorized in this, concurrent, and future replies, to charge payment for any additional fees that may be required, or credit any overpayment, to Deposit Account No. 50-0238.

Respectfully submitted,

VOLENTINE FRANCOS, P.L.L.C.



Andrew J. Telesz, Jr.
Reg. No. 33,581

12200 Sunrise Valley Drive, Suite 150
Reston, VA 20191
Tel. No. (703) 715-0870
Fax No. (703) 715-0877

Enclosures: Appendix A
Appendix B

Claims - Appendix A

6. A semiconductor device comprising:

first and second gates formed on active regions of a substrate, said first and second gates each consisting of a refractory metal layer on a polysilicon layer;

a field oxide formed on the substrate between said first and second gates;

side walls formed on side surfaces of said first and second gates, said side walls being a silicon oxide film;

a protective layer formed selectively on said field oxide to prevent overetching of said field oxide, said protective layer being a material different than said field oxide; and

an insulating layer, a contact hole, and a connecting wire formed above a surface of the substrate.

7. The semiconductor device of claim 6, wherein said protective layer is a polysilicon layer.

8. The semiconductor device of claim 6, wherein said protective layer is formed on said field oxide only.

9. The semiconductor device of claim 6, wherein said first and second gates are MOSFET gates.

11. A semiconductor device comprising:

a gate formed on an active region of a substrate;

a field oxide formed on the substrate adjacent the active region;

a protective layer formed on said field oxide to prevent overetching of said field oxide, said protective layer being a material different than said field oxide; and

an insulating layer, a contact hole, and a connecting wire formed above a surface of the substrate,

said protective layer being formed on said field oxide only.

12. The semiconductor device of claim 11, wherein said protective layer is a polysilicon layer.

13. The semiconductor device of claim 11, wherein said gate is a MOSFET gate.

14. The semiconductor device of claim 11, further comprising side walls formed on said surfaces of said gate, said side walls being covered by said insulating layer.

15. The semiconductor device of claim 11, further comprising an additional gate formed on the substrate, said field oxide being formed on the substrate between said gate and said additional gate.

16. A semiconductor device comprising:

a gate formed on an active region of a substrate, said gate consisting of a refractory metal layer on a polysilicon layer;

side walls formed on side surfaces of said gate, said side walls being a silicon oxide film;

a field oxide formed on the substrate adjacent the active region;

a protective layer formed on said field oxide to prevent overetching of said field oxide, said protective layer being a material different than said field oxide; and

an insulating layer, a contact hole, and a connecting wire formed above a surface of the substrate,

said protective layer being formed on said field oxide only.

17. The semiconductor device of claim 16, wherein said protective layer is a polysilicon layer.

18. The semiconductor device of claim 16, wherein said gate is a MOSFET gate.

19. The semiconductor device of claim 16, further comprising an additional gate formed on the substrate, said field oxide being formed on the substrate between said gate and said additional gate.

SILICON PROCESSING

FOR

THE VLSI ERA

VOLUME 2:

PROCESS INTEGRATION

STANLEY WOLF Ph.D.
Professor, Department of Electrical Engineering
California State University, Long Beach
Long Beach, California

LATTICE PRESS

Sunset Beach, California

DISCLAIMER

This publication is based on sources and information believed to be reliable, but the authors and Lattice Press disclaim any warranty or liability based on or relating to the contents of this publication.

Published by:

Lattice Press,
Post Office Box 340
Sunset Beach, California 90742, U.S.A.

Cover design by Roy Montibon, Visionary Art Resources, Inc., Santa Ana, CA.

Library of Congress Cataloging in Publication Data
Wolf, Stanley

Silicon Processing for the VLSI Era
Volume 1 : Process Integration

Includes Index
1. Integrated circuits-Very large scale integration. 2. Silicon. I. Title

86-081923

ISBN 0-961672-4-5

9 8 7 6

PRINTED IN THE UNITED STATES OF AMERICA

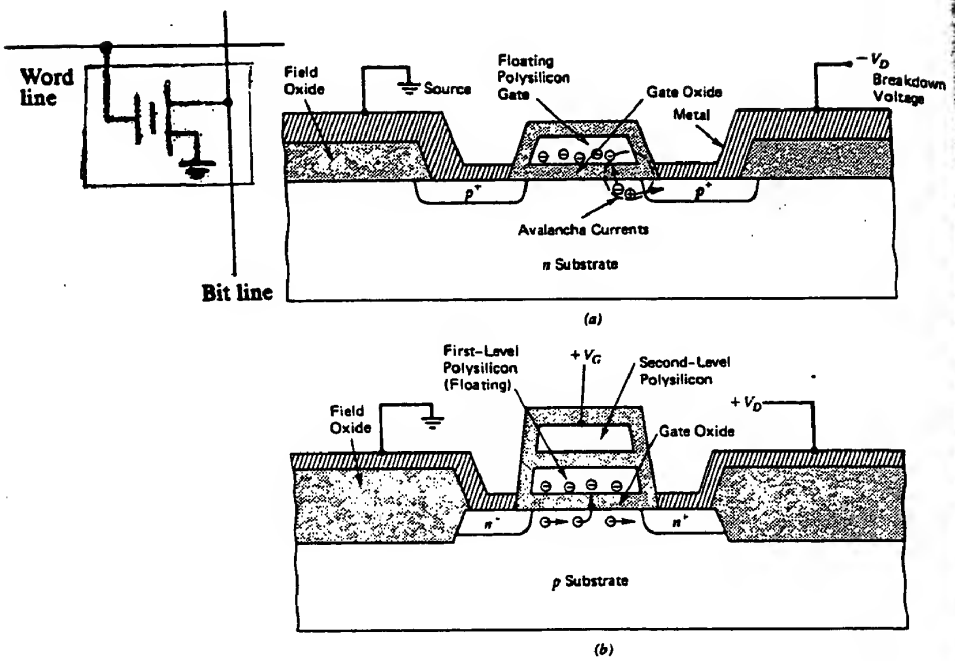


Fig. 8-33 (a) Circuit schematic and cross section showing the mechanism of charge injection into the gate by avalanche in a FAMOS memory element. (b) A FAMOS element made with two layers of polysilicon and suitable for *n*-channel MOS applications. From R. S. Muller and T. I. Kamins, *Device Electronics for Integrated Circuits,* 2nd Ed. Copyright 1986, John Wiley & Sons. Reprinted with permission.

Traditionally, such EPROMs have been used as prototyping vehicles to ensure that no glitches remained in the code. Once the programs were finalized, the code was usually fixed into ROM components. However, the cost of EPROMs is shrinking with advances in technology, and as a result, their use is growing at the expense of ROMs. Another factor in favor of EPROMs is their faster turn around time (which also plays a role in the choice of technology used to implement masked ROMs).

The charge-transfer mechanism is based on the injection of hot electrons into the floating polysilicon gate, which is completely encapsulated by SiO_2. The original EPROM devices were fabricated in PMOS technology and consisted simply of a MOSFET with a floating gate (Fig. 8-33a). If a sufficiently high reverse-bias voltage is applied to the drain, the drain-substrate *pn* junction will experience avalanche breakdown, causing hot electrons to be generated. Some of these will have enough energy to pass over the oxide potential-energy barrier and charge the floating gate (see section 5.6.2). These EPROM devices were thus called ***Floating-gate, Avalanche-injection MOS transistors*** (FAMOS).

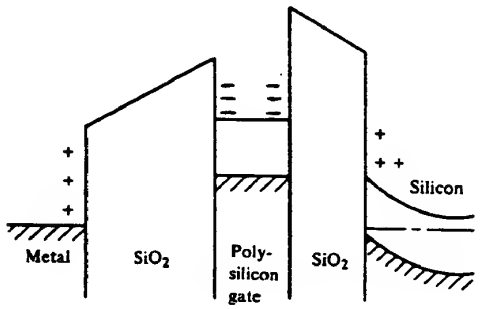


Fig. 8-34 Energy band diagram of a FAMOS device with charge stored in the silicon gate. From E. S. Yang, *Microelectronic Devices*, Copyright 1988, McGraw-Hill Book Company. Reprinted with permission.

Once electrons are transferred to the gate, they are trapped there, as illustrated by the energy-band diagram shown in Fig. 8-34. Since the potential-energy barrier at the oxide-silicon interface is greater than 3 eV, the rate of spontaneous emission of electrons from the oxide over this barrier is negligibly small. The electronic charge on the floating gate can thus be retained for many years.

If the floating gate is charged with a sufficient number of electrons, inversion of the channel under the gate will occur. A conducting channel then forms between the source and the drain, exactly as would occur if an external gate voltage were applied. The presence of a *1* or *0* in each bit location is therefore determined by the presence or absence of a conducting channel in a programmed device.

Subsequent advances in process technology (Fig. 8-33b) made it possible to implement EPROMs with 5 V, *n*-channel devices.[83,84] In such EPROMs the cells can also be laid out in NOR or NAND arrays; we will use the NOR array configuration to describe the operation of these newer cells.

Two layers of polysilicon are used to form a double gate in the transistor, as shown in Fig. 8-33b. Gate #1 is the floating gate and is placed under Gate #2. Cell selection is controlled by Gate #2, which therefore plays the role of the single gate in conventional MOS transistors. Initially, Gate #1 is uncharged; thus, if the drain, source, and Gate #2 of the transistor are grounded, Gate #1 will also be at 0 V. If a voltage (V_2) is subsequently applied to Gate #2, the voltage on Gate #1 (V_1) will be given by:

$$V_1 = [C_2 / (C_1 + C_2)]\ V_2 \qquad (8-3)$$

because the two gates represent a capacitive divider as shown in Fig. 8-35. From the electrical perspective of Gate #2, the transistor appears to have a larger V_T. In order to turn on this transistor, a larger gate voltage must be applied to Gate #2 (typically somewhat more than twice the normal V_T). For example, if a conventional NMOS

window these MOS PROMs can be field-programmed only once. As a result, they are commonly known as *one-time-programmable* (OTP) ROMs. The advantage of these over bipolar PROMs is that they can be fabricated in much higher densities. High-density CMOS OTP ROMs are now being built with access times close to those of bipolar PROMs, but with more bits per chip and much lower power dissipation. For example, a 256-kbit CMOS OTP ROM with an access times of 50 ns has recently been introduced; this approaches the access time (~40 ns) of large [64-kbit] bipolar PROMs.[87] In addition, OTP ROMs are no longer much more expensive than ROMs, and hence they are also expected to increasingly replace masked ROMs.

8.6 ERASABLE PROGRAMMABLE READ-ONLY MEMORIES (EPROMS)

Erasable PROMs depend on the long-term retention of electronic charge as the information-storage mechanism. The charge is stored on a *floating polysilicon gate* of an MOS device (the term *floating* refers to the fact that no electrical connection exists to this gate). The charge is transferred from the silicon substrate through an insulator.

Each of the various mechanisms implemented to transfer (and remove) charge from the floating gate has been the basis of a different erasable-PROM device type. This section describes the so-called *electrically programmable ROM* (EPROM), which also requires that the device be irradiated with ultraviolet (UV) light for removing (or *erasing*) the stored charge from the floating gate.

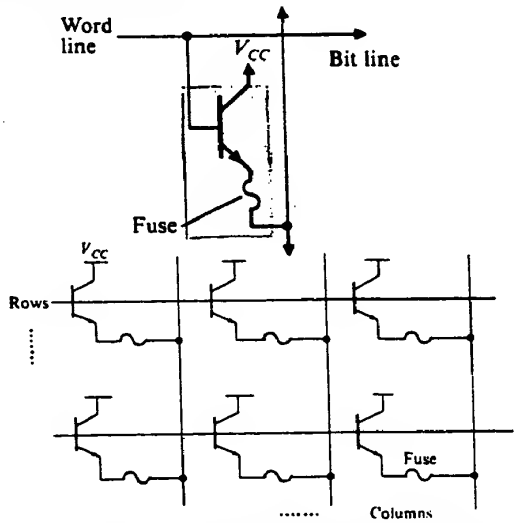


Fig. 8-32 (a) Emitter-follower bipolar PROM.[1] From D. A. Hodges and H. G. Jackson, *Analysis and Design of Digital Integrated Circuits*, Copyright, 1983 McGraw-Hill Book Co. Reprinted with permission.

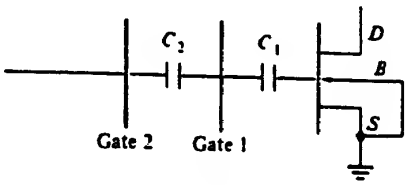


Fig. 8-35 Equivalent capacitive divider of an EPROM structure.

device with a $V_T = 1$ V was fabricated with the double gate of Fig. 8-33b, a voltage of 2 V would have to be applied to Gate #2 to turn it *ON*; a voltage of 5 V for reading the cell would also cause it to turn *ON* (Fig. 8-36). Such a turned-on device would cause a *positive logic stored zero* to appear at the output of the bit line if the device was used in a NOR array. As a result, the programming of the EPROM begins by discharging all of the floating gates through exposure to UV radiation, so that every cell initially stores a *0*. A *1* is then selectively written into the desired cells.

For a *1* to be written into a cell, both Gate #2 and the drain are raised to about 12 V (for a few hundred microseconds), while the source and substrate are kept grounded (early EPROMs required 30 V programming voltages for several milliseconds). Hot electrons are created near the drain and are attracted to the floating gate (which, due to capacitive coupling, has a more positive potential than the drain). Some fraction of the electrons will traverse the oxide and charge the floating gate. When the voltages on Gate #2 and the drain are returned to zero, these charges remain trapped on Gate #1. The electrons trapped on Gate #1 cause its potential to be at about -5 V. Therefore, if a signal of only 5 V is applied to Gate #2 when the EPROM is being read, no channel will form in the transistor. Under this circumstance, a *1* is stored in the cell. The electron- trapping process is self-limiting, because once electrons are stored on the floating gate they begin to inhibit further electron injection.

In order for the cells to be erased, the stored charge must be removed from the floating gate. This is accomplished by flood exposure of the EPROM with strong ultraviolet light for approximately 20 minutes. The UV light creates electron-hole pairs in the SiO_2, providing a discharge path for the charged floating gate.

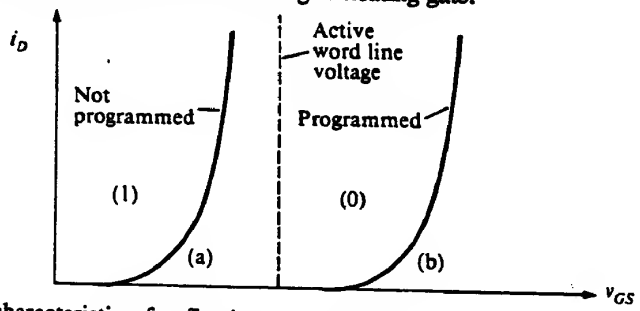


Fig. 8-36 Transfer characteristic of a floating-gate transistor.

One of the advantages of EPROMs is that the cells consist of only one transistor, allowing them to be fabricated with high densities (e.g., a 4-Mbit CMOS EPROM with an access time of 120 ns and 0.8-μm channel-length transistors has been reported).[85] In addition, they cost less to manufacture than electrically erasable PROMs (EEPROMs – see the next section).

A disadvantage of EPROMs is that they require UV light for erasing and must therefore be packaged in an expensive ceramic package with a UV-transparent quartz window. In addition, they must be removed from the circuit board and put into a special UV eraser. (Note that since sunlight and fluorescent lamps contain some UV, one week of sunlight or three years of room-level fluorescent lighting are likely to erase some of the cells. Therefore, except during erasure, the window should be covered at all times with an opaque label.) Another disadvantage is that the high voltage needed to program the EPROM is generally not available on the integrated circuit, so a special programming setup must also be provided. This limitation, combined with the fact that EPROM programming takes a relatively long time, means that these cells are used primarily for reading information and are only occasionally rewritten. (Note, however, that the programming time is decreasing dramatically. In the first 64-kbit EPROMs, it took about 50 ms to program each byte, adding up to almost seven minutes for the entire chip; in the 4-Mbit EPROM,[85] the program time has been reduced to 10 μs/byte, so that the entire chip can be programmed in only five seconds!)

OTP ROMs compete with high-density masked ROMs because they offer the benefit of a significantly shorter TAT (albeit at a somewhat higher cost). OTP ROMs are also less expensive than bipolar PROMs, and offer a PROM capability with much higher density. While bipolar PROMs are generally faster, a three-transistor EPROM cell was recently reported that would allow CMOS EPROMs to be built with the same speed and density as bipolar PROMs, but with much lower power dissipation and 100% testability.[86] Another one-transistor cell, split-gate 256-kbit CMOS EPROM with an access time of 50 ns has also been reported.[87]

Some of the relevant process and circuit-design enhancements used in fabricating the large CMOS EPROMs include the following:[88]

- Use of thin (20 nm) reliable interpoly dielectric materials, often consisting of composite films of $SiO_2/Si_3N_4/SiO_2$, for increased capacitive coupling between Gates #1 and #2.
- Self-aligned contacts to the control gate (as well as a self-aligned floating gate) to achieve the 3.1 x 2.9 μm^2 small cell size.
- Use of a low-resistance polycide gate for the word line to achieve high speed.
- Reduction of the programming voltage to 10.5 V, along with a reduction of the programming time to ~10 μs/byte.
- On-chip test circuits.

A novel self-aligned planar-array EPROM cell has also been proposed.[89,90] This

cell appears to make possible the fabrication of 4-Mbit EPROMs with 1-μm design rules because it uses buried n^+ bit lines that are self-aligned to the FAMOS transistor.

8.7 ELECTRICALLY ERASABLE PROMS (EEPROMS)

In some applications it is desirable to erase the contents of a ROM electrically, rather than to use a UV light source. In other circumstances it is useful to be able to change one byte at a time, without having to erase the entire IC. A variety of *electrically erasable PROMs* have been developed to serve these applications. Such EEPROMs are the most sophisticated of the ROM families in terms of the physical operating principles and process complexity. For example, EEPROMs must be fabricated with unique tunnel oxides, as well as with high-voltage transistors (for programming and erasing the devices).

Three technologies have been developed for EEPROM fabrication: (1) MNOS transistors; (2) Floating-gate Tunnel Oxide (FLOTOX) MOS transistors; and (3) textured-polysilicon floating-gate MOS transistors. Although MNOS transistor-based devices were among the first EEPROMs to be commercially manufactured, their technology limitations have made them less widely adopted than the others. Therefore, we will devote most of our attention to FLOTOX and textured-poly EEPROMs.

8.7.1 MNOS-Based EEPROMs

The MNOS EEPROM cell consists of a single MOS-like transistor that employs a composite gate-dielectric layer (Si_3N_4, ~50 nm thick, on top of a very thin [~2 nm thick] SiO_2 layer), as shown in Fig. 8-37. (See ref. 118 for more details on MNOS devices.) Unlike in floating-gate MOS devices, the charge is stored in discrete traps in the nitride bulk. The charge transfers from the substrate to the nitride traps (and back, during erasure) by tunneling through the thin oxide layer. Programming is accomplished by applying a high voltage to the top gate; erasing is done by grounding the top gate and raising the well to a high potential. MNOS transistors are built within wells (akin

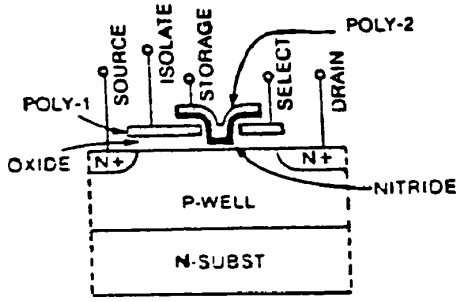


Fig. 8-37. Cross-sectional structure of an MNOS memory cell.[107] (© 1983 IEEE).

to those used in CMOS) so that their channel potentials can be controlled.

The manufacturing process of MNOS transistors involves the following modifications to the standard single polysilicon-gate MOS technology: thin oxide growth, nitride deposition, and post-nitride temperature cycles. Mastering the processes used to grow the ultra-thin oxide and deposit the nitride and to control their quality is a challenging task. Furthermore, while the basic transistor is very small and highly scalable, each cell of the memory array requires a select transistor. This requirement, coupled with the need to fabricate wells, produces a relatively large effective cell size. Finally, the charge stored on the nitride traps continually leaks away through the thin oxide by means of tunneling, even when no erase voltage is applied. The charge loss is thus time dependent, making charge retention the main reliability concern with MNOS devices. (With MNOS structures, as the switching speed is increased, the ability to retain stored charge is decreased. Thus, devices with a retention time of tens of years can be fabricated if a slow switching speed can be tolerated.)

Nevertheless, MNOS exhibits higher tolerance to ionizing radiation than do either of the other EEPROM technologies. Thus, MNOS EEPROMs currently find their main use in low-density military applications that need radiation-hardened EEPROMs; this appears to be the niche to which MNOS EEPROMs will be relegated in the future.[91]

8.7.2 FLOTOX EEPROMs

The *floating-gate tunneling oxide (FLOTOX)* transistor, shown in Fig. 8-38a, consists of an MOS transistor with two polysilicon gates. A thin (8-12 nm) gate oxide (or oxynitride) region is formed near the drain. The lower polysilicon layer is the floating-gate while the other is the control gate. The remainder of the floating-gate oxide is typically 50 nm thick, and the interpoly oxide is ~50 nm thick. Programming of this transistor is done by causing electrons to be transferred from the substrate to the floating

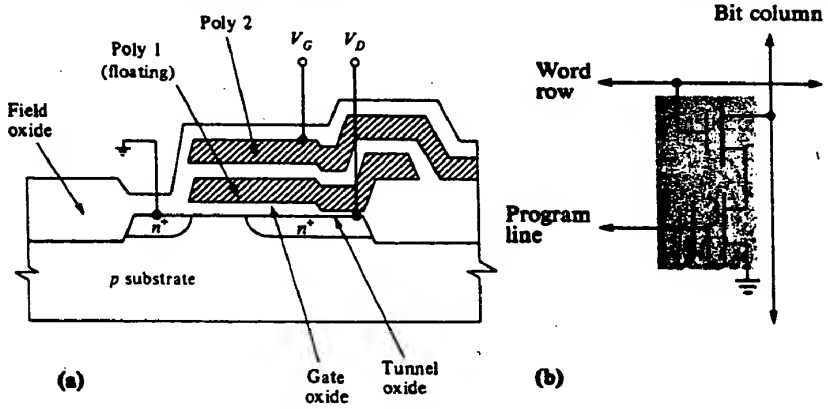


Fig. 8-38 (a) Cell structure of a Flotox transistor structure, (b) Connection in an EEPROM.[91] (© 1986 IEEE).

gate through the thin oxide layer by means of Fowler-Nordheim tunneling.[92]

The control-gate voltage is raised to a sufficiently high value so that tunneling ensues (e.g., 12 V in modern FLOTOX EEPROMs). As electronic charge builds up on the floating gate, the electric field is reduced, which decreases the electron flow. Since the tunneling process is reversible, the floating gate can be erased by grounding the control gate and raising the drain voltage, indicating that tunneling is used both to *program* and *erase* the FLOTOX transistor. Programming and erase times are on the order of 9 ms. Electron transfer by Fowler-Nordheim tunneling, however, requires a minimum electric- field strength of around 10 MV/cm. Thus, for oxides of 10 nm in thickness, such tunneling will be negligible when normal 5-V signals are applied. As a result, FLOTOX transistors can be expected to retain their charges for more than 10 years if the memory is subjected only to normal read cycles.

The FLOTOX transistor must be isolated by a select transistor. Otherwise, the high voltage applied to the drain of the selected cell during erasing would also appear on the drain of the other *unselected* cells in the same memory column. A FLOTOX EEPROM cell must therefore consist of two transistors (Fig. 8-38b). Although this limits the density of such EEPROMs in comparison to EPROMs and flash EEPROMs, it makes it possible to erase and re-program one byte of the memory without having to erase the entire IC. In addition, two cycles are needed to load the correct data into the memory. In the first, all the cells in a byte are programmed (i.e., the floating gates are charged); in the second, selected cells are erased, with the drain used for data control.

The fabrication of FLOTOX EEPROMs involves a modification of the polysilicon-gate MOS process. A double-polysilicon process is used, together with a thin tunnel-oxide growth process. The growth of a high-quality, thin tunneling oxide is, in fact, the critical manufacturing step in this technology. The tunneling dielectric reportedly can be successfully implemented with nitrided oxides, since the barrier between silicon nitride and silicon is lower than that between SiO_2 and Si. As a result, a higher tunneling current can be obtained for the same voltage.[93]

Despite the fact that a reliable process for growing thin tunneling oxides must be developed, FLOTOX-based devices have become the most widely manufactured of the EEPROM types. They are still the easiest to learn to manufacture for companies that have already successfully developed an EPROM process. Since it is desirable to be able to program and erase the EEPROM while it remains in place on a PC board, considerable effort has also been expended to make this memory type fully operational with a 5 V power supply. (This type of operational capability is referred to as *5 V only*.)

FLOTOX-based EEPROMs are best suited for applications in which low-cost, low-density, nonvolatile memories are required – for example, in microcontrollers and programmable logic devices. Another potential application is for smart credit cards; several Japanese companies have announced 64-kbit FLOTOX-based EEPROMs for this market.

On the other hand, scaling and reliability considerations appear to limit the maximum size of FLOTOX EEPROMs to 256 kbits. The need for two transistors, and the relatively large size of the select transistor (due to the large voltages needed for

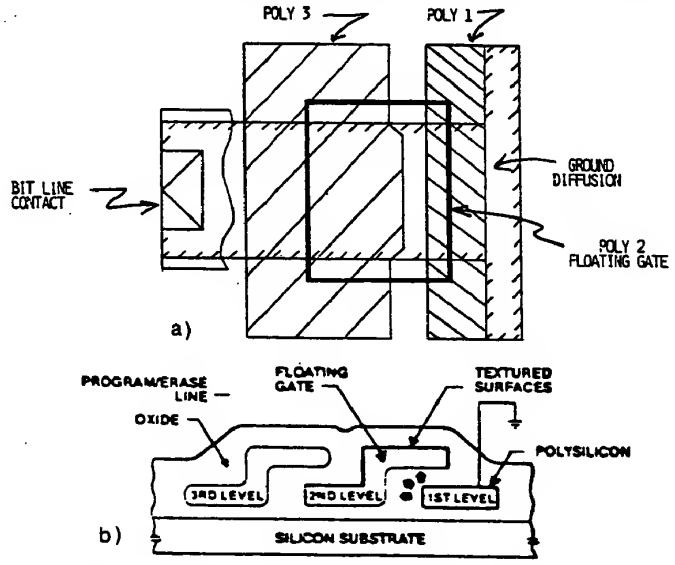


Fig. 8-39 Textured-polysilicon memory cell. (a) Top view. (b) Cross-sectional structure.[91] (© 1986 IEEE).

programming and erasure) are contributing factors. The poly-to-poly area of the sense transistor must also be large, due to the high oxide capacitance of the thin tunnel oxide.[40] Furthermore, the FLOTOX EEPROMs exhibit high failure rates, caused by defect-related oxide-breakdown problems as memory size is increased.[94] Error-detection and correction codes (EDCC) can be used to overcome this limitation, but such solutions impose a penalty of increased die cost. Reference 95 gives an example of a 50-ns access time, 256-kbit FLOTOX-based CMOS EEPROM using a single-bit EDCC.

The reliability of FLOTOX-based EEPROMs compares favorably with that of the other two EEPROM types. As with the others, there is a very low failure rate during 5-V operation; reliability problems occur as a result of the high voltages that must be used during programming and erasing. Random single-bit failures occur in FLOTOX devices due to oxide defects, resulting in leaky oxides that lose charge over time. The number of cycles that most FLOTOX EEPROMs are specified to be able to endure before the thin oxide becomes too leaky to retain data sufficiently, is 10^3–10^4 cycles. However, a process for increasing this endurance level to 10^6 cycles has been reported.[96]

8.7.3 Textured-Polysilicon EEPROMs

Textured-polysilicon EEPROMs, introduced in 1983 as an alternative to the tunneling oxide types of devices, are also based on the floating-gate MOS technology. The cell consists of three layers of polysilicon that partially overlap (Fig. 8-39) to create a cell

that behaves like three MOS transistors in series. The floating-gate MOS transistor is formed by the middle polysilicon structure, which is encapsulated with SiO_2 to enable high charge retention. While charge is still transferred to the floating gate by means of Fowler-Nordheim tunneling, tunneling takes place from one polysilicon structure to another rather than from the substrate to the floating gate. The interpoly oxides through which the tunneling takes place can be made significantly thicker than the tunneling oxides in FLOTOX devices (60-100 nm in textured poly devices, versus <12 nm in FLOTOX devices), since the electric field that promotes the tunneling is enhanced by the geometrical effects of the fine texture at the surface of the polysilicon structures.

Textured-poly cells are programmed by causing electrons to tunnel from *poly 1* to the *floating poly*. Erasure is accomplished by causing electrons to tunnel from the floating poly structure to *poly 3*. The voltage of poly 3 is taken high in both the programming and erase operations. The drain voltage, however, determines whether tunneling occurs from *poly 1* to the *floating gate*, or from the *floating gate* to *poly 3*. As a result, the state of the drain voltage determines the final state of the memory cell. This provides an advantage, in that the cell represents a *direct write cell* – there is no need to charge all the cells and then remove the charge from selected cells, as with FLOTOX EEPROMs.

Textured-poly EEPROMs depend on a tunneling process whose physical mechanisms are not as well understood as those of tunneling through thin oxides, and which appears to require tighter control of empirically determined process parameters. In addition, the three poly layers require a more complex (and therefore more costly) fabrication sequence. Furthermore, textured-poly EEPROMs require a higher operating voltage than FLOTOX devices (>20 V). Finally, an intrinsic endurance problem is caused by the very high electron trapping that occurs as a result of tunneling in the poly oxides. This eventually leads to a condition in which the memory can no longer be programmed or erased.

For all of the above reasons, the textured-poly approach has been less widely pursued than the FLOTOX approach. Only one company, Xicor, is heavily involved in manufacturing these devices.[97] However, because the poly cells can be made about one-half the size of FLOTOX cells, it is possible to fabricate them in high-density configurations. In 1989, the largest textured-poly EEPROMs being offered had a 1-Mbit capacity. Although the cell-size advantage gives the textured-poly approach an edge over the FLOTOX EEPROMs for memories larger than 256 kbits, the flash-EEPROM technology described in the next section, provides a way to achieve equally high-density EEPROMs without the need to develop a textured-poly process.

8.8 FLASH EEPROMS

The *flash EEPROM* device is so named because the contents of all of the memory's array cells can be erased simultaneously as with a UV-EPROM, but through the use of an electrical erase signal. The term *flash* refers to the fact that the cells can be erased much more rapidly (1 or 2 seconds, compared to the 20 minutes required to erase a UV-EPROM). Although it was not possible to erase only a single byte in the first

generation of flash EEPROMs, by 1989 parts had become available that offered a byte-by-byte erasable (and 64-byte erasable) feature in a 256-kbit memory.[98]

Flash EEPROMs are attractive for the middle of the programmable semiconductor spectrum, where neither EPROMs nor EEPROMs are particularly cost effective. The applications in this range typically require more memory capacity than EEPROMs can provide, but they also need faster and more frequent reprogramming than can be accomplished with EPROMs. Examples include automotive and automated factory equipment applications. As an example, the average EPROM cost about $7 in 1989, and a flash memory about $25. But the differential is wiped out by the expense of single reprogramming. The in-system reprogramming of a flash device may cost as little as $1, whereas pulling an EPROM out of a system to erase it by exposure to 20 minutes of UV light may cost over $80 when equipment, downtime and labor are factored in.

Meanwhile, EEPROMs are likely to remain popular wherever bytes will have to be erased selectively. But flash products, might do better for updating stored logic, when this must done more than once but less often than in main memory, cache memory, or registers. Reprogramming costs are similar, but flash memories are less than half the price of EEPROMs.

The erasing mechanism in flash EEPROMs is Fowler-Nordheim tunneling off the floating gate to the drain region. Programming of the floating gates, however, is carried out in most flash cells by *hot-electron injection into the gate.** Unlike floating-gate EEPROMs (which incorporate a separate select transistor in each cell to allow individual byte erasure), flash memories forego the select transistor to obtain bulk erasure. Thus, flash-EEPROM cells are roughly two to three times smaller than floating-gate EEPROM cells fabricated with the same design rules.[99] Figure 8-40 shows the cross-section of a CMOS flash-EEPROM cell implemented with triple polysilicon, and a SEM photo of a double-poly flash EEPROM cell.

Most flash-EEPROM cells use a double-poly structure, as shown in Fig. 8-41 (which also shows the Toshiba triple-poly cell, Fig. 8-41b). The upper poly forms the control gate and the word lines of the structure, while the lower poly is the floating gate. The gate oxide is ~10 nm thick,[100] and the interpoly dielectric is an oxide/nitride/oxide composite film ~45 nm thick.[99] In the structure shown in Fig. 8-40 and 8-41c the control-gate poly overlaps the channel region adjacent to the channel under the floating gate. This structure is needed because when the cell is erased, it leaves a positive charge on the floating gate. As a result, the channel under the floating gate becomes inverted. The series enhancement-mode transistor (formed by the control gate over the channel region), is needed in order to prevent current flow from source to drain. A more recently reported flash-EEPROM cell (Fig. 8-41a) does not require the control gate to form a series enhancement-mode transistor, because it uses a special software-controlled erase procedure that prevents the floating gate from being over erased.[100]

* The 5-V-only flash memories from Texas Instruments and Amtel depend on tunneling for both write and erase mechanisms.

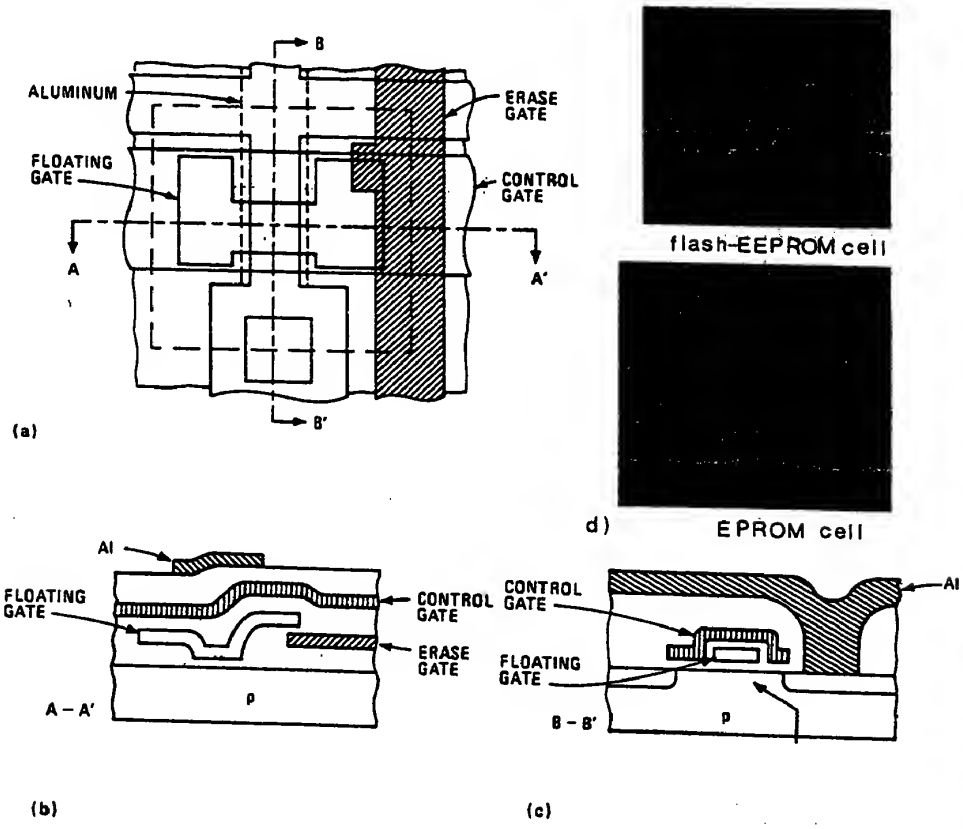


Fig. 8-40 Triple-poly flash-EEPROM cell from Toshiba: (a) Layout; (b) Cross section of the cell; (c) Section at right angles to the section shown in part (c);[101] (d) SEM pictures of double-poly flash-EEPROM and EPROM cells.[99] (© 1988 IEEE).

Flash EEPROMs can be seen to combine the advantages of UV-erasable EPROMs and floating-gate EEPROMs. They offer the high density (Fig. 8-42), small-die size, lower cost, and hot-electron writability of EPROMs, together with the easy erasability, on-board reprogrammability, and electron-tunneling erasure features of EEPROMs. High-density CMOS flash EEPROMs in 1-Mbit sizes are commercially available. It is projected that by the year 2000, 256-Mbit flash EEPROMs will be fabricated with 0.25 μm geometry.

With a memory-cell size of about one-quarter the size of current EEPROM cells, the flash EEPROMs also achieve EPROM die sizes. In addition, there are two types of flash EEPROMs: (1) those that are more akin to the EEPROM (and thus require a 12-V

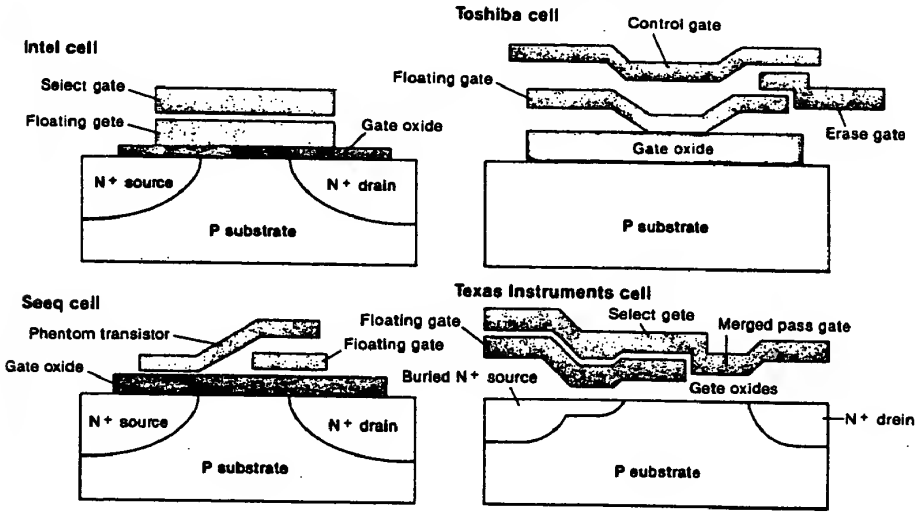


Fig. 8-41 Four approaches to flash memory technology: (a) Intel cell; (b) Toshiba triple-polysilicon cell; (c) SEEQ cell; (4) Texas Instruments 5-V-only cell.[125] (© 1989 IEEE).

external supply for programming and erasure); and (2) those that are closer to EPROMs, and hence need only a 5-V supply. Furthermore, the programming voltage can be applied during read operations, eliminating the need to switch it off when not erasing or programming. Byte-write times are 100 μs, and erasure times are 200 ms. Access times of 110 ns at 30-mA active-current consumption are provided by a 128-kBit CMOS flash EEPROM.[100,101] Endurance (i.e., the number of times a device can be erased and written) is a minimum of 100 cycles, and can be as high as 1000 cycles (note that this is lower than the endurance of EEPROMs, which is typically 1000 – 10,000 cycles).

8.9 NONVOLATILE FERROELECTRIC MOS RAMS

A novel type of nonvolatile MOSFET DRAM memory cell, introduced in late 1987, uses the electrical polarization of a ferroelectric capacitor to store information semipermanently.[102,103] Since ferroelectric polarization retention is nearly perpetual (just as in magnetic core memories), refresh is not needed. The reported write speed is 200 ns in one design,[102] and 60 ns in another,[103] which is much faster than that exhibited by an EEPROM (1 ms) or a UV-PROM (10 ms) without fatigue after 10^{12} write cycles. It is predicted that by 1991 products will be available with operating lifetimes of ~75 years at a cycle time of 100 ns, and 10^{12} read/write cycles. A recent review article has described the latest advances in such nonvolatile RAMs.[124]

IEEE ELECTRON DEVICE LETTERS, VOL. 10, NO. 3, MARCH 1989

Degradations Due to Hole Trapping in Flash Memory Cells

SAMEER HADDAD, CHI CHANG, BALAJI SWAMINATHAN, AND JIH LIEN

Abstract—Degradation in the hot-electron programmability of the flash memory cell is observed after erasing from the drain. Trapped holes in the oxide near the drain junction are found to be responsible for this degradation. Hole trapping in the oxide also causes another problem known as "gate disturb," which is the undesired increase in the threshold voltage of an erased cell during programming of the other cells on the same word line. Threshold-voltage shifts due to gate disturb are used to monitor the amount of trapped holes in the oxide after cell erasure. It is determined that the trapped holes are mainly externally injected from the junction depletion region rather than directly generated in the oxide by the Fowler–Nordheim (F–N) tunneling process.

FLASH memory [1]–[3] has recently emerged as an important nonvolatile memory which combines the advantages of EPROM density with EEPROM electrical erasability. Programming of the cell is achieved by hot-electron injection at the drain side, and erasure is usually accomplished by electron tunneling from the floating gate to either the drain [2] or the source [1], [3]. However, due to the presence of high voltage at the gated-diode junction during erasure, holes are inevitably generated by band-to-band tunneling [4] and a small amount of them are injected into the oxide after being accelerated in the depletion region. Hot-hole injection during erasure has been reported to cause variations in the erased threshold voltages of the cells in the memory array [1], and trapped holes in the oxide were shown to degrade the charge-retention characteristics of the memory cells [5]. However, no work has been reported on the effect of hole trapping on the programmability of the flash cell when erasing is done at the drain. In this work this effect is studied and the degradation in programmability is characterized. Since hole trapping in oxide is known to alter the tunneling characteristics of the oxide [6], [7], we have devised a method to compare hole trapping produced by Fowler–Nordheim (F–N) tunneling with that produced by hot-hole injection. Results clearly demonstrate that oxide hole trapping in flash structures is indeed attributable to the externally injected hot holes.

The flash cell used in this work has a structure very similar to that of a conventional stacked-gate EPROM cell. The channel length is about 1 μm long and the S/D arsenic diffusion is about 0.25 μm deep. The gate oxide used is less than 120 Å thick and the erasure mechanism is dominated by F–N tunneling [7]. In Fig. 1(a) three dc programming curves are presented for a fresh cell, for the cell electrically erased from the source, and for another cell electrically erased from the drain. It shows significant degradation in the programmability of the cell that has been erased from the drain side. Fig. 1(b) shows the programmed and erased threshold voltages as a function of the number of program/erase (P/E) cycles for a cell erased from the source and another erased from the drain. Fig. 1(b) clearly shows the programming threshold voltage is reduced after erasure from the drain, and as the cell is cycled the programming window continues to close. No degradation in the programmability is observed after erasure from the source. We believe that the degradation in programmability is caused by hole trapping in the oxide.

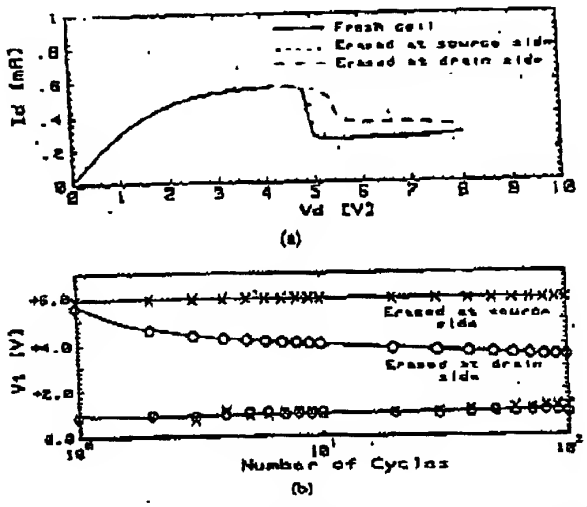


Fig. 1. (a) DC programming curves for (V_g = 12 V) a fresh cell (solid), electrically erased from the source side (dotted), and electrically erased from the drain side (dashed). (b) Programming/erase threshold voltage of the flash memory cell versus number of cycles, for drain-side and source-side erasure.

To demonstrate hole trapping in the oxide as a result of high-voltage erasure, poly gate transistors with similar gate oxide thickness and junction profile to the flash cell were used to monitor the hot-electron-generated substrate current and the gated-diode leakage current before and after the high-voltage stress. The stress was performed by applying a constant voltage of 11.5 V to the drain junction for 10 s while the gate and substrate were grounded with the source left floating. Fig. 2(a) shows the substrate current as a function of the gate voltage for V_D = 4.V, before and after the stress. The substrate current becomes lower after the stress due to a

Manuscript received October 21, 1988; revised December 19, 1988
The authors are with Advanced Micro Devices, Inc., Sunnyvale, CA 94088.
IEEE Log Number 8926657.

0741-3106/89/0300-0117$01.00 © 1989 IEEE

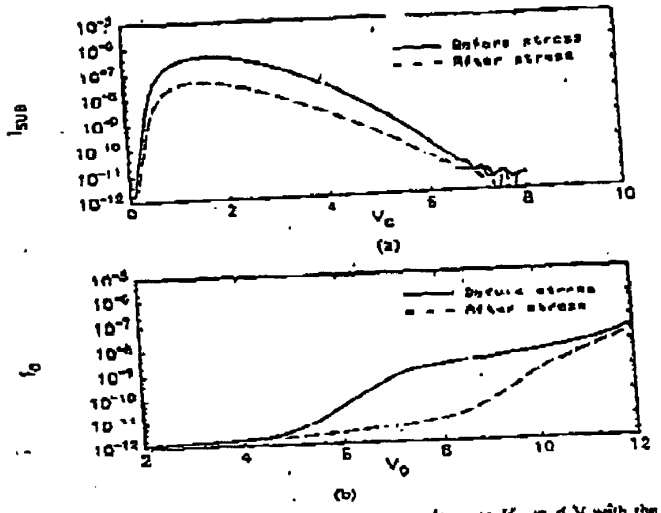


Fig. 2. (a) Substrate current versus the gate voltage at $V_D = 4$ V with the source grounded. (b) Drain leakage current versus drain voltage at $V_G = 0$ V with the source floating, measured before (solid) and after (dashed) high-voltage stress ($V_D = 11.5$ V for 10 s).

reduction in the channel field, indicating hole trapping in the oxide near the drain junction [8]. Fig. 2(b) shows the band-to-band tunneling current as a function of the drain voltage before and after stress. The drain leakage current versus V_D curve after the stress is shifted to a much higher drain voltage, also suggesting the fact that the surface field at the junction corner has been reduced due to hole trapping [4]. Hole trapping in the oxide near the drain has the effect of reducing the maximum channel electric field during programming, thereby decreasing channel hot-electron generation. This explains the degradation in programmability after drain-side erase.

By erasing from the source, programmability degradation is eliminated and, furthermore, the drain and source junctions can be independently optimized. The drain junction can be made shallow and abrupt to enhance the hot-electron effect for programming, and the source junction can be graded (for example, using a double-diffused junction) to reduce the junction field during erasing. A flash cell using this concept, which was first proposed by Kume *et al.* [1], is used in the following study to determine the origin of the trapped holes in the oxide during the erase cycle.

It is known that hole trapping inside the oxide reduces the barrier for electron tunneling [6]. This barrier lowering effect will lead to the so-called "gate disturb" problem for a memory array. Gate disturb refers to the undesired increase in the threshold voltage of the unselected "erased" cell during programming of the other cells on the same word line. During programming, a fraction of the word-line voltage will be coupled to the floating gate, and a high field (about 6–7 MV/cm) will appear across the thin gate oxide. As a result, electrons can tunnel to the floating gate through the thin oxide, causing the threshold voltage to increase. The gate disturb at a given field is proportional to the F–N tunneling current which is enhanced by the hole trapping in the oxide. Therefore gate disturb is used in this work as a monitor of the amount of trapped holes in the oxide due to erasure.

We believe that holes were introduced into the oxide through hot-hole injection during flash erasure. However, another possible mechanism is hole generation by impact ionization in the oxide during F–N tunneling itself [9], similar to that experienced in EEPROM erasure. The gate disturb test was used to compare positive charge trapping in the oxide using flash erasure with that using EEPROM-type erasure. For EEPROM-type erasure a negative voltage pulse (−16 V) is applied to the control gate, while the source junction is kept at a low positive voltage (2 V). The source voltage is low enough not to generate hot holes, while electron tunneling still occurs predominantly at the gate–source overlap region. On the other hand, the flash erasure is exercised by applying a high positive voltage pulse (11 V) to the source with the gate grounded and the drain floating. Fig. 3 shows the threshold voltage of the flash cell versus the erase time using the two different erase conditions mentioned above. For both erasures, the erasing speed is controlled by F–N tunneling. The voltages were chosen such that the oxide field during the EEPROM-type erasure is slightly larger than that for the flash erasure. This is done to ensure that the F–N tunneling-induced hole generation in the oxide for the EEPROM-type erasure is no less than that in the flash erasure.

Fig. 4 compares the gate disturb after EEPROM and flash erasures. The cell was first alternately programmed and erased to V_t's of 6 and 1 V, respectively, for 20 cycles in order to saturate the hole trapping in the oxide. Then the threshold voltage was measured as a function of gate disturb time. During the gate disturb, the gate of the erased cell was biased at 12 V while the other terminals were at ground. As shown in the figure, no V_t shift is observed for the cell erased under EEPROM condition. This suggests that the holes that are

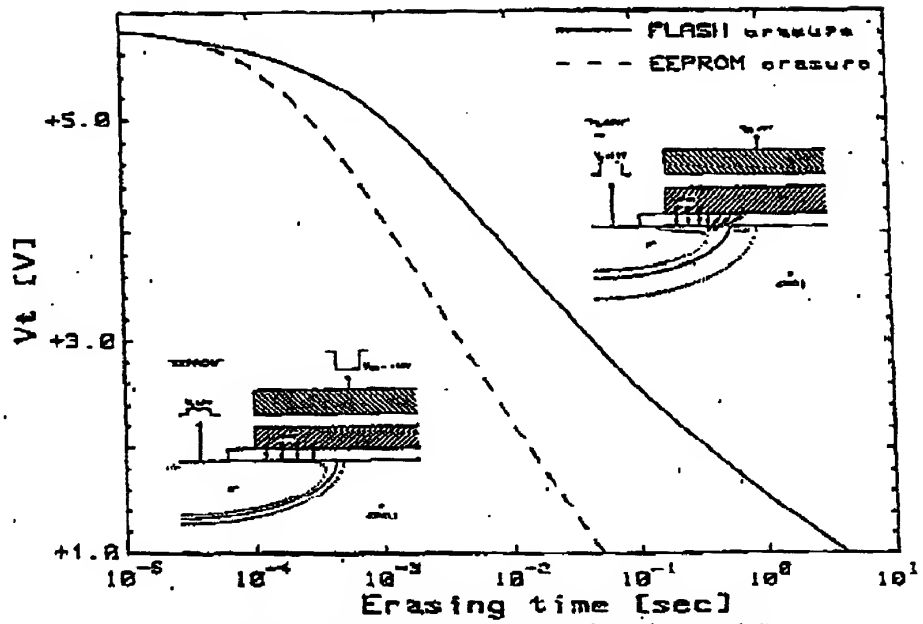


Fig. 3. Threshold voltage versus erasing time during flash erasure (solid) and EEPROM-type erasure (dashed). The faster erasure time of the EEPROM-type erasure is due to the higher applied field across the oxide. The insets are schematic representations of the two erase conditions.

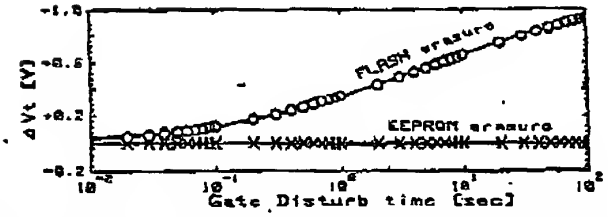


Fig. 4. Threshold-voltage shifts ($\Delta V_t = V_t - V_t$(before disturb)) during gate disturb after the cells having been cycled 20 times, using flash (○) and EEPROM (×) erasure. The cells were erased to a V_t of 1 V before gate disturb measurement.

generated and trapped in the oxide by the F–N tunneling electrons are insignificant in contributing to gate disturb. However, the gate disturb after the flash erasure is significant. These data strongly support a model of external hot-hole injection and hole trapping in the oxide during the flash erasure.

In summary, we have demonstrated that the hot-electron programming efficiency of the flash cell is degraded after erasure from the drain. Degradation in programmability has been shown to be caused by trapped holes in the gate oxide at the drain junction. By comparing the flash erasure to EEPROM-type erasure, it has been determined that these trapped holes are externally injected from the junction depletion region. Furthermore, the trapped holes in the gate oxide will lead to gate disturb. Therefore, in a properly designed flash structure, it is essential that hot-hole injection is minimized.

References

[1] H. Kume *et al.*, "A flash-erase EEPROM cell with an asymmetric source and drain structure," in *IEDM Tech. Dig.*, 1987, pp. 560–563.

[2] G. Samachisa *et al.*, "A 128K flash EEPROM using double-polysilicon technology," *IEEE J. Solid-State Circuits*, vol. SC-22, pp. 676–683, Oct. 1987.

[3] S. Mukherjee, T. Chang, R. Pang, M. Knecht, and D. Hu, "A single transistor EEPROM cell and its implementation in a 512K CMOS EEPROM," in *IEDM Tech. Dig.*, 1985, pp. 616–619.

[4] C. Chang and J. Lien, "Corner-field induced drain leakage in thin oxide MOSFET's," in *IEDM Tech. Dig.*, 1987, pp. 714–717.

[5] G. Verma and N. Mielke, "Reliability performance of ETOX based flash memories," in *26th Proc. IEEE Reliability Phys. Symp.*, Apr. 1988, pp. 158–166.

[6] I. C. Chen, S. Holland, and C. Hu, "Electrical breakdown of thin gate and tunneling oxides," *IEEE Trans. Electron Devices*, vol. ED-32, no. 2, pp. 413–422, 1985.

[7] C. Chang, S. Haddad, B. Swaminathan, and J. Lien, "Drain avalanche and hole-trapping induced gate leakage in thin-oxide MOS devices," *IEEE Electron Device Lett.*, vol. 9, no. 11, pp. 588–590, 1988.

[8] Y. Nissan-Cohen, G. Franz, and R. Kwasnick, "Measurement and analysis of hot-carrier-stress effect on NMOSFET's using substrate current characterization," *IEEE Electron Device Lett.*, vol. EDL-8, pp. 451–453, 1987.

[9] I. C. Chen, S. Holland, and C. Hu, "Hole trapping and breakdown in thin SiO_2," *IEEE Electron Device Lett.*, vol. EDL-7, no. 3, pp. 164–166, 1986.